スパースな分散表現に基づく Web サイトの関係分析モデルに関する研究

情報数理応用研究　5218C032-4　保坂大樹
指導教員　後藤正幸

A Study on Analytical Model of Web Sites Relationship Based on Sparse Embedding

HOSAKA Taiju

1. 研究背景・目的

近年，インターネット技術の発達と PC やスマートフォンの普及に伴い，インターネット上のビジネスが盛んになっている．また，各企業は広告配信などの Web マーケティング施策を講じることで，効率的に新規顧客を獲得しようとしている．ここで，Web マーケティングを行う企業において，消費者の嗜好を検出する方法は重要な課題の一つである．例えば，検索駆動型広告では，消費者の検索語に紐づき，関連した内容の広告が配信される．また，成果報酬型広告では，配信者が設定した Web サイトにアクセスした消費者を広告の配信対象とする．さらに，消費者のデモグラフィック属性ごとの閲覧傾向を分析することで，デモグラフィック属性に応じてどのような施策を講じるべきかを論じた研究 [1] もある．

ここで，消費者の閲覧行動には各消費者の嗜好が反映されると考えられるため，Web 閲覧履歴データ（以下，閲覧履歴データ）は重要なマーケティング分析の対象となっており，多額のコストを投じて収集している企業もある．すなわち，各消費者の嗜好は閲覧した Web サイトの特徴により説明できることが多い．例えば，自動車関係の Web サイトをよく閲覧する消費者は，自動車に興味があると考えられる．

そこで本研究では，より多角的なマーケティング分析を行うために，閲覧履歴データをもとに，各 Web サイトの特徴を推定するためのモデルを構築することを考える．これにより，ある消費者の閲覧に関して，類似した内容の Web サイトを連続して閲覧している場合には，特定の内容に興味を持っていると解釈できる．逆に，閲覧している Web サイトの特徴が頻繁に変化する場合には，特定の閲覧目的を持たずに閲覧をしていることが想定される．このことから，連続して閲覧されている Web サイト間の関係を分析することで,「自動車」などの特定のトピックに興味を持つきっかけとなる Web サイトを発見したり，閲覧履歴系列中に内在する消費者の興味を検出し，その内容に応じた広告を配信することが可能になると考えられる．

ここで，自然言語処理分野において，Word2Vec[3] と呼ばれる分散表現モデルが，単語の意味解析で高い性能を示し注目されている．Word2Vec は，学習コーパス中で生起する単語に対して，前後に生起する単語との類似性を仮定し，オンライン学習法に基づいて単語の意味を考慮したベクトル表現を学習する．類似した意味を持つ単語には同程度のベクトル表現が与えられ，このベクトル表現を一般的に分散表現と呼ぶ．これに対して，インターネット上の閲覧行動では，長期間では様々な閲覧目的のもとで多様な Web サイトを閲覧するが，短期間では特定の閲覧目的のもとで類似した内容の Web サイトを閲覧することが多い．したがって，前後のデータに類似性を仮定する Word2Vec の特性は閲覧履歴データに対して自然に拡張することができる．実際，閲覧履歴データ分析に対する Word2Vec の有効性が報告されている [2]．

一方で，Word2Vec が対象とする言語データと閲覧履歴データの間には大きな特性の違いも存在する．言語データ中の単語が文脈中の明確な意味に基づいて生起することに対して，閲覧履歴データ中の閲覧行動がその閲覧目的を反映しない場合も考えられる．例えば，同時に複数の閲覧目的を持ちながら閲覧を行う場合には，前後に閲覧された Web サイトの閲覧目的が異なる可能性が考えられる．また,「閲覧途中に友人から電話を受けた」などの外的なイベントにより，突然に閲覧目的が変化する可能性もある．このように，閲覧履歴データ中には，前後に閲覧されていたとしても Web サイト間に類似性を仮定するべきではない閲覧行動も存在する．

本研究では，一貫した目的に基づいた閲覧系列は閲覧履歴データ中で何度も繰り返し生起する一方で，閲覧目的が一貫していない閲覧系列はほぼ繰り返し生起しないことに着目する．そして，閲覧履歴データ中に存在する Web サイト間の強い関係性のみを抽出し，ノイズの影響を抑制することを目的として，スパース正則化を導入した分散表現の学習アルゴリズムを提案する．これにより，ノイズとなる閲覧の影響を排除し，Web サイト間の本質的な関係性がモデル化され，意味空間上で Web サイトの総合的な関係分析が可能となる．最後に，提案モデルを実際の閲覧履歴データに適用し，得られた表現に基づく分析を行い，提案モデルの有効性を検証する．

2. 準備

2.1. 分散表現

自然言語処理の分野において，データを構成する単語をどのように数値表現するべきであるかは，大きな課題の一つである．これに対して，近年，単語の意味を考慮した分散表現の重要性が高まっている．分散表現では，分析者があらかじめ設定した次元数のベクトルで各単語を表現する．そのため，適切なアルゴリズムにより，予め与えられた文書コーパスから，類似した意味を持つ単語の組には互いに距離が近いベクトル表現が与えられるように学習する．すなわち，学習が成功した分散表現モデルでは，ベクトル間の距離により，単語間の意味的な類似性を定量的に測定することが可能である．

代表的な分散表現モデルの一つに，Mikolov らによって提案された Word2Vec[3] がある．Word2Vec では，コーパス中に生起する単語は周辺の単語から予測が可能であるという仮説のもとで，ニューラルネットワークの構造に基づい

て分散表現を推定する．Word2Vec における分散表現の学習モデルの一つである Skip Gram モデルでは，学習コーパス中の単語の One-hot ベクトルを入力として，周辺の単語を予測する多クラス分類問題を解く 2 層のネットワークを学習し，各入力ノードから中間層への学習済みの重みベクトルが各単語の分散表現として抽出される．一方，Word2Vec におけるもう一つの学習モデルである CBoW モデルでは，周辺の単語から中心の単語を予測するネットワークを学習し，中間層から各出力ノードへの学習済みの重みベクトルが各単語の分散表現として抽出される．また，各モデルの損失の計算には，多クラス分類問題を少数の二クラス分類問題で近似する階層ソフトマックスや負例サンプリングなどの計算量削減技法が広く用いられている．

2.2. 関連研究

本研究では，Web サイトの関係分析モデルを，スパースな分散表現で構築することを目的としている．自然言語処理の分野においても，解釈性の観点から，スパースな分散表現を推定する方法について，いくつかの研究が行われている．

Faruqui ら [4] は，行列分解モデルに正則化を考慮したオンライン学習法を導入して，学習済みの単語の分散表現を，比較的高次元な潜在意味空間上のスパースなベクトルに線形写像する手法を提案した．潜在意味空間の次元数を高くしながらスパース化を行うことで，表現力を保ちつつ解釈性に優れたベクトルの獲得が期待できる．実際，この研究では，潜在意味空間の各軸の解釈性を評価する Word Intrusion Detection Test を行い，提案手法によりベクトルの解釈性が向上していることが示されている．一方，Sun ら [5] は，正則化を考慮したオンライン学習法である Regularized Dual Averaging[8] を CBoW モデルに導入し，スパースな単語の分散表現を学習する手法を提案した．この研究では，先述の関連研究と異なり，スパース化を行いながら各単語の分散表現を直接推定するアルゴリズムを構築している．

3. 提案モデル

3.1. 概要

Web サイトの特徴表現を獲得するために，分散表現モデルの適用は有効である [2]．しかしながら，閲覧履歴データ中には，閲覧目的が一貫していない閲覧行動が存在する．このような閲覧行動により，異なる閲覧目的のもとで閲覧された Web サイトに類似性を仮定してしまう可能性が考えられる．

本研究では，分散表現モデルに L1 正則化を加えることで，閲覧履歴データ中のノイズとなる関係性の抽出を制御する分析モデルを提案する．具体的には，正則化を考慮したオンライン学習アルゴリズムである FTRL-Proximal[7] により Skip Gram モデルの学習アルゴリズムを改良し，Web サイトのスパースな分散表現を推定する．これまで，類似したモデルとして，Sun らが CBoW モデルに Regularized Dual Averaging を適用したアルゴリズムを提案している．しかしながら，CBoW モデルは文脈から単語を予測するために，同一の文脈のもとで予測単語の競合が発生し，低頻度語の推定精度が不安定になる．また，McMahan ら [7] によって提案された FTRL-Proximal は，過去のパラメータの値を考慮しながらパラメータを更新するため，過去の勾配のみを考慮する Regularized Dual Averaging よりも精度良くパラメータを推定することが可能であることが知られている．

3.2. 定式化

閲覧履歴データ中のユーザ数を M，ユーザ x_m $(1 \leq m \leq M)$ の閲覧セッション集合を $\mathcal{Y}_m$，閲覧セッション $y \in \mathcal{Y}_m$ 内の閲覧数を I_y とし，Web サイト数を V とする．閲覧履歴系列中の Web サイトに対して類似性を仮定する前後の幅をウィンドウサイズと定義し，W で表す．また，各 Web サイトに対してサンプリングする負例の数を負例サンプルサイズと定義し，K で表す．ここで，Web サイト v $(1 \leq v \leq V)$ に対して D 次元の Web サイトベクトル $\boldsymbol{u}_v$，および文脈ベクトル $\boldsymbol{c}_v$ を定義する．また，Web サイト v の全閲覧履歴データ中の生起頻度を $freq_v$ として，式 (1) で雑音分布 $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_v, \ldots, \theta_V)^\top$ を構築する．

$$\theta_v = \frac{(freq_v)^{3/4}}{\sum_{v'=1}^{V} (freq_{v'})^{3/4}} \tag{1}$$

提案モデルでは，ユーザ x_m の閲覧セッション $y \in \mathcal{Y}_m$ 内の閲覧履歴系列中の Web サイト w_i^y とその前後閲覧サイト w_{i+j}^y $(-W \leq j \leq W,\ j \neq 0)$ に対して，それぞれ K 個の負例サイト w_k $(1 \leq k \leq K)$ が $\boldsymbol{\theta}$ からサンプリングされ，式 (2) で損失 $L(w_{i+j}^y | w_i^y)$ が計算される．

$$\begin{aligned} &L(w_{i+j}^y | w_i^y) = \\ &-\log \sigma(\boldsymbol{c}_{w_{i+j}^y}^\top \boldsymbol{u}_{w_i^y}) - \frac{1}{K} \sum_{k=1}^{K} \log \sigma\left(-\boldsymbol{c}_{w_k}^\top \boldsymbol{u}_{w_i^y}\right) \end{aligned} \tag{2}$$

ここで，$\sigma(\cdot)$ は標準シグモイド関数である．また，Web サイトベクトルに対する L1 正則化の重みを λ とすると，閲覧履歴データ全体に対する損失 L は式 (3) で計算される．

$$\begin{aligned} L = \sum_{m=1}^{M} \sum_{y \in \mathcal{Y}_m} \sum_{i=1}^{I_y} \sum_{-W \leq j \leq W, j \neq 0} L(w_{i+j}^y | w_i^y) \\ + \lambda \sum_{v=1}^{V} \|\boldsymbol{u}_v\|_1 \end{aligned} \tag{3}$$

ここで，$\|\boldsymbol{a}\|_1$ はベクトル $\boldsymbol{a}$ の L1 ノルムを表す．

3.3. パラメータの学習アルゴリズム

Web サイトベクトル $\boldsymbol{u}_v$ の t 回目の更新時の値を $\boldsymbol{u}_{v,t}$ と表記する．また，Web サイトベクトル $\boldsymbol{u}_v$ の t 回目の更新時の勾配の和を $\boldsymbol{g}_{v,t}$，勾配の二乗和を $\boldsymbol{g}_{v,t}^2$ で表す．

ここで，$\boldsymbol{u}_{v,t-1}$ の学習率 $\eta_{v,t}$ は Adagrad[8] に基づき，式 (4) で計算される．ただし，η を学習率の初期値とし，ε を調整パラメータとする．

$$\eta_{v,t} = \frac{\eta}{\varepsilon + \sqrt{\boldsymbol{g}_{v,t}^2}} \tag{4}$$

最後に，$\boldsymbol{u}_{v,t}$ の d 番目の要素 $u_{v,t,d}$ $(1 \leq d \leq D)$ は式 (5)〜(7) に基づいて値が計算される．

$$\phi_{v,t} = \begin{cases} \frac{1}{\eta_{v,t}} - \frac{1}{\eta_{v,t-1}} & t \geq 2 \\ \frac{1}{\eta_{v,t}} & t = 1 \end{cases} \tag{5}$$

$$\boldsymbol{z}_{v,t} = \eta_{v,t}\left(\sum_{\tau=1}^{t}\phi_{v,\tau}\boldsymbol{u}_{v,\tau-1} - \boldsymbol{g}_{v,t}\right) \tag{6}$$

$$u_{v,t,d} = \begin{cases} 0 & |z_{v,t,d}| \le \lambda \\ z_{v,t,d} - \mathrm{sign}(z_{v,t,d})\lambda & \text{otherwise} \end{cases} \tag{7}$$

文脈ベクトル $\boldsymbol{c}_v$ $(1 \le v \le V)$ については，Skip Gram と同様に，確率的勾配降下法を用いてパラメータ推定を行う．

4. 実データ分析

本研究では，提案モデルの有効性を検証するために，実際の閲覧履歴データに提案モデルを適用し，スパースな Web サイトの分散表現を獲得する．また，獲得した分散表現を用いて，ユーザの閲覧行動に関する分析を行い，考察を与える．

4.1. 分析条件

モデルの学習には，株式会社ヴァリューズ提供の Web 閲覧履歴データを用いる．このデータは，登録に同意したモニタが PC またはスマートフォンから閲覧した Web サイトのホスト名を記録したものである．ただし，ホスト名で Web サイトを区別しているため，同じタイトルを持つページが複数存在する場合がある．対象データの収集期間は，2017 年 8 月 1 日から 2017 年 10 月 31 日，総閲覧数は 78, 508, 580 回，ユーザ数は $M = 49,787$ 人，Web サイト数は $V = 47,854$ 個である．以下では，モデルのパラメータについて述べる．

各 Web サイトに対応づける Web サイトベクトル，文脈ベクトルの次元数をそれぞれ $D = 200$ とし，ウィンドウサイズを $W = 4$，負例サンプルサイズを $K = 20$ とした．また，L1 正則化の重みを $\lambda = 2.5$，Web サイトベクトルと文脈ベクトルの学習率の初期値を $\eta = \alpha = 0.5$ とした．さらに，Adagrad の調整パラメータを $\varepsilon = 1.0$ とした．

4.2. 関連研究との比較

本節では，提案モデルにより得られる表現が関連研究で得られる表現と比較してどの程度スパースであるかを検証する．

ここで，閲覧履歴データ中のノイズとなる関係性の抽出をどれだけ抑制しているかをを示す指標として，Web サイトベクトルの平均スパース率 S_{vec}，Cos 類似度行列の平均スパース率 S_{sim} をそれぞれ式 (8), (9) で定義する．ただし，$\mathbb{1}(\cdot)$ は括弧内の関係が成立すれば 1, しなければ 0 を返す指示関数であり，$\mathrm{Cos}(\boldsymbol{a}, \boldsymbol{b})$ は $\boldsymbol{a}$ と $\boldsymbol{b}$ の Cos 類似度を表す．

$$S_{\mathrm{vec}} = \frac{1}{VD}\sum_{v=1}^{V}\sum_{d=1}^{D}\mathbb{1}(u_{v,d} = 0) \tag{8}$$

$$S_{\mathrm{sim}} = \frac{2}{V(V-1)}\sum_{v=1}^{V-1}\sum_{v'=v+1}^{V}\mathbb{1}(\mathrm{Cos}(\boldsymbol{u}_v, \boldsymbol{u}_{v'}) = 0) \tag{9}$$

各モデルの Web サイトベクトルの平均スパース率 S_{vec} および Cos 類似度行列の平均スパース率 S_{sim} を表 1 に示す． Mikolov らの手法では，スパース性を考慮した学習を行っていないため，Web サイトベクトルの要素は 0 にはならず，密なベクトルを構成することが表 1 からも確認できる．一方，Faruqui らの手法や Sun らの手法では，調整したパラメータにより，Web サイトベクトルのスパース率が 90% を超えている．しかしながら，これらの手法では Web サイトベクトル間の Cos 類似度はほとんど 0 にならないことがわかる．Faruqui らの手法では，学習済みの分散表現を，その関係性を維持しながらスパースな表現に変換する．すなわち，この手法では，Mikolov らの手法によって構成された密な関係性をスパース化することができない．また，Sun らの手法で用いられている CBoW モデルは，文脈ベクトルを平均化し，文脈全体で Web サイトとの類似性を仮定しているため，被閲覧数の大きい Web サイトに依存した特徴を学習する可能性を指摘することができる．これに対し，Skip Gram モデルに基づく提案モデルでは，Web サイトベクトルをスパース化するだけでなく，Cos 類似度行列もスパース化できていることがわかる．この結果は，Web サイトが意味空間上の複数の部分空間に分布する構造をなすことを示唆している．

表 1: 各モデルのスパース性

	S_{vec}	S_{sim}
Mikolov ら [3]	0.0%	0.0%
Faruqui ら [4]	93.2%	0.0%
Sun ら [5]	94.7%	1.0%
提案モデル	95.2%	22.1%

4.3. 提案モデルを用いた分析

本節では，提案モデルを用いることで可能になる分析の例を提示する．あるユーザの閲覧系列および直前に閲覧した W サイトとの平均 sign-dot 類似度を図 1 に示す．ただし，D 次元ベクトル $\boldsymbol{a}, \boldsymbol{b}$ 間の sign-dot 類似度は，どれだけの軸を共有しているかを表した類似度として，式 (10) で定義する．

$$\text{sign-dot}(\boldsymbol{a}, \boldsymbol{b}) = \sum_{d=1}^{D}\mathrm{sign}(a_d)\mathrm{sign}(b_d) \tag{10}$$

図 1 より，このユーザは「Yahoo!天気・災害」といった天気予報サイト、「日刊スポーツ」などのニュースサイト、「Gmail」などの Web ツールや「シネマトゥデイ」といった映画情報サイトを閲覧したのちに自動車関係の Web サイトを複数閲覧し，最後に「みずほ銀行」といった金融機関の Web サイトを閲覧している．このとき，9 番目の「webCG」から 16 番目の「Car Watch」までは，直前の Web サイトとの類似度が群として高くなっている部分閲覧系列である．このような閲覧系列は意味空間上の同一の部分空間内における閲覧行動であり，自動車関係の Web サイトへの興味が高まっていることを表現している．これに対して，直前の Web サイトとの類似度が低い部分閲覧系列は，意味空間上の部分空間が切り替わりながら，すなわち閲覧目的が変わりながら閲覧が行われていることが想定され，ネットサーフィンのような閲覧行動であると解釈することができる．このように，実際の閲覧行動に，Web サイトベクトルにより測定される Web サイトの類似度を重ね合わせることで，閲覧行動におけるユーザの興味を検出することができると考えられる．また，提案モデルは Web サイト間の類似性の強弱を強調した

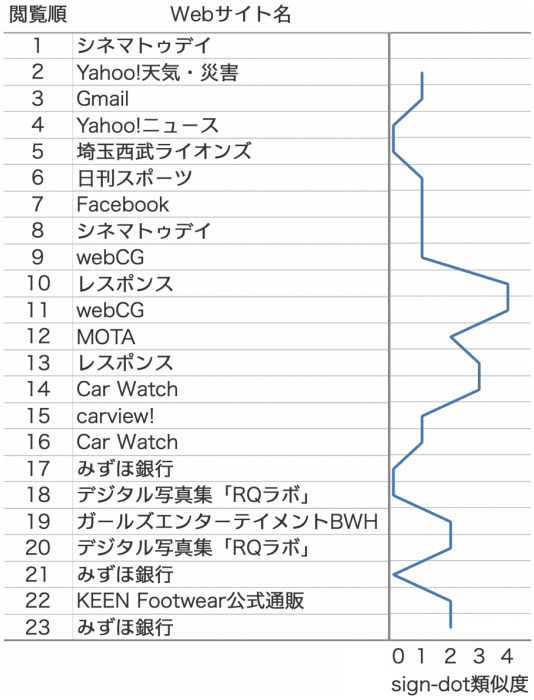

図 1: 消費者の閲覧系列と直前の Web サイトとの類似度

表現を学習するために，本節の分析のような Web サイト間の強い関係性と弱い関係性を総合した分析に有効である．

5. 考察

従来の関係データに関する研究の多くは，オブジェクト間の強い関係を抽出するためのモデルを構築する．例えば，EC サイトでは，顧客の購買履歴データに基づいて，購買された商品と類似した商品を各顧客に推薦する．しかしながら，複数のオブジェクト間の関係を包括的に扱う場合，強い関係を抽出するだけでは，弱い関係とノイズとなる関係が混在してしまう．したがって，このような状況においては，提案モデルのように，強い関係を抽出するだけでなく，ノイズとなる関係を抑制することが重要であると考えられる．

4.3 節の分析では，ユーザの閲覧セッション内で Web サイト間の関係を分析している．直前の Web サイトとの類似度が高くなるような閲覧は，特定の内容への興味が高まっていることを示す．図 1 では，ユーザの自動車関係の Web サイトに対する興味を検出したことから，このユーザに対して，自動車関係の広告を配信することが有効であると考えられる．ここで，ユーザの閲覧の状態は閲覧と同時に推定することが可能であるため，リアルタイムの広告配信に活用することができる．また，自動車関係の Web サイトへの興味が検出された部分閲覧系列は，「webCG」を閲覧することから始まっている．このことから，「webCG」を閲覧することは，他の自動車関係の Web サイトの閲覧を促進する可能性がある．そこで，他の閲覧系列と統合して分析を行い，「webCG」が自動車関係の Web サイトの閲覧を促すと結論づけられた場合，自動車関係の Web サイトの運営者は，自社サイトに新規顧客を誘導するために，自社サイト上に「webCG」との相互リンクを作成することが望ましいといえる．

6. まとめと今後の課題

本研究では，多角的にマーケティング分析を行うことを目的として，分散表現モデルにオンライン正則化学習法を導入することで，閲覧履歴データ中に存在する Web サイトのノイズとなる関係の抽出を制御し，意味空間上で Web サイト間の関係性を総合的に分析可能な関係分析モデルを構築した．

実データへの適用では，提案モデルは従来の分散表現モデルと同様に Web サイトの特徴を学習する一方で，重要でない Web サイト間の類似度を 0 としていることが明らかになった．また，Web サイトの類似度の変動の観点から消費者の閲覧行動を分析し，提案モデルの有用性を示した．

今後の課題として，よりマクロな視点からの分析への応用が挙げられる．例えば，Web サイト間の類似度が 0 になりやすい提案モデルは，同時に多数の Web サイトの関係を扱うネットワーク分析との相性が良いと考えられる．また，従来の分散表現モデルが対象としている言語データにおいて，提案モデルの有効性を確認することも重要な課題の一つである．

謝辞

本研究にあたり，熱心な議論と貴重なデータの提供をいただいた株式会社ヴァリューズの皆様に深く感謝いたします．

参考文献

[1] I. Weber and C. Castillo, “The Demographics on Web Search,” *Proc. the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.523–530, 2010.

[2] Y. Tagami, H. Kobayashi, S. Ono and A. Tajima, “Representation Learning for Users’ Web Browsing Sequences,” *IEICE Transactions on Information and Systems*, Vol. E101.D, No. 7, pp.1870–1879, 2013.

[3] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, “Distributed Representations of Words and Phrases and their Compositionality,” *Adv. the 26th Conference on Neural Information Processing Systems*, pp.3111–3119, 2013.

[4] M. Faruqui, Y. Tsvetkov, D. Yogatama, C. Dyer and N. A. Smith, “Sparse Overcomplete Word Vector Representations,” *Proc. the 53rd Annual Meeting of the Association for Computational Linguistics*, pp.1491–1500, 2015.

[5] F. Sun, J. Guo, Y. Lan, J. Xu and X. Cheng, “Sparse Word Embeddings Using $\ell 1$ Regularized Online Learning,” *Proc. the 25th International Joint Conference on Artificial Intelligence*, pp.2915–2921, 2016.

[6] L. Xiao, “Dual Averaging Methods for Regularized Stochastic Learning and Online Optimization,” *Journal of Machine Learning Research*, Vol. 11, pp.2543–2596, 2010.

[7] H. B. McMahan, G. Holt, D. Sculley, M. Young, D. Ebner, J. Grady, L. Nie, T. Phillips, E. Davydov, D. Golovin, S. Chikkerur, D. Liu, M. Wattenberg, A. M. Hrafnkelsson, T. Boulos and J. Kubica, “Ad Click Prediction: a View from the Trenches,” *Proc. the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.1222–1230, 2013.

[8] J. Duchi, E. Hazan and Y. Singer, “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization,” *Journal of Machine Learning Research*, Vol. 12, pp.2121–2159, 2011.